2018

Effects of early language experiences on the auditory brainstem

BOSTON UNIVERSITY

SARGENT COLLEGE OF HEALTH AND REHABILITATION SCIENCES

Thesis

EFFECTS OF EARLY LANGUAGE EXPERIENCES ON THE AUDITORY BRAINSTEM

by

ANDREA C. CHANG

B.Sc., University of Alberta, 2016

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2018

Approved by

First Reader ______________________________

Tyler K. Perrachione, Ph.D.
Assistant Professor of Speech, Language, and Hearing Sciences

Second Reader ______________________________

Melanie L. Matthies, Ph.D., CCC-A
Senior Associate Sargent College Dean
Associate Professor and Chair of Speech, Language, and Hearing Sciences

Third Reader ______________________________

Barbara G. Shinn-Cunningham, Ph.D.
Professor of Biomedical Engineering
Boston University, College of Engineering

Fourth Reader ______________________________

Sung-Joo Lim, Ph.D.
Postdoctoral Fellow of Speech, Language, and Hearing Sciences

DEDICATION

I would like to dedicate this work to my friends and family who have provided endless support to me throughout this journey.

ACKNOWLEDGMENTS

Over the past two years, my classmates in the Boston University Masters of Speech-Language Pathology program have provided me immense support and encouragement. I am highly appreciative of my committee members, Dr. Melanie Matthies, Dr. Barbara Shinn-Cunningham, and Dr. Sung-Joo Lim, who have generously given their time, insight, and expertise. My fellow lab members were imperative to the completion of this project. Thank you to Yaminah Carter and Michelle Njoroge for helping run this experiment. Thank you to Deirdre McLaughlin, Terri Scott, Ja Young Choi, Jayden Lee, and Sara Beach for your expertise regarding speech science and help regarding this project. A huge thank you to Jessica Tin for her coding expertise and help throughout this project. A special thank you to Dr. Le Wang for his dedication to this project, for answering my many questions, for helping me design the Frequency Following Response paradigm, and for guiding me with the analysis. Finally, thank you Dr. Tyler Perrachione, for encouraging me to complete a thesis through your passion for science, for guiding me and supporting me throughout this journey, for the countless hours spent reading and revising the text, and for your great insight and dedication to this project.

EFFECTS OF EARLY LANGUAGE EXPERIENCES

ON THE AUDITORY BRAINSTEM

ANDREA C. CHANG

ABSTRACT

Recent studies have come to contradicting conclusions as to whether international adoptees, who experience a sudden change in language environment, lose or retain traces of their birth language (Pallier et al., 2003; Ventureyra, Pallier & Yoo, 2004; Pierce, Klein, Chen, Delcenserie, & Genesee, 2014). Though these studies have considered cortical differences between international adoptees and individuals from their birth counties, none has looked at subcortical differences in the brain between the two groups. The current project examined the frequency following response of adult Chinese international adoptees (N = 9) adopted as infants by American English-speaking families in the United States compared to native Mandarin (N = 21) and American English (N = 21) controls. Additional behavioral tasks were completed to explore different levels of linguistic features from phonetics to phonology to semantic knowledge to suprasegmental characteristics of speech. The FFR results indicate mostly good pitch tracking abilities amongst the adoptees that may support future tonal language learning in the adoptees. The behavioral data suggest that the adoptees have minimal access to all levels of linguistic levels of linguistic processing (i.e., phonetic, phonological, lexical, suprasegmental) after adoption and after early exposure to English. Overall, the data provide evidence for the neural commitment theory that humans' language acquisition is attuned to their language environment early on in life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

During the first few months of life, all infants with normal hearing perceive most of the phonological distinctions in all human languages (Eimas et al., 1971; Kuhl, 2004). However, to acquire a language, it is necessary to learn the phonetic distinctions used in that language. With maturation, an infant's language environment reduces the phonetic differences they can perceive (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984). Kuhl et al. argue that this reduction in foreign-language perception is due to "neural commitment" to acoustic parameters specific to an infant's linguistic environment (Kuhl, Tsao, Zhang, & De Boer, 2001). As a result, this "neural commitment" interferes with the acquisition of phonological patterns that do not conform to an individual's specific linguistic environment. The ontogeny of lexical tone perception, one such phonological contrast used in tone languages like Mandarin Chinese, is inconclusive, with some suggesting that acquisition occurs before one years of age (Mattock & Burnham, 2006; Mattock, Molnar, Polka & Burnham, 2008; Yeung, Chen & Werker, 2013), whereas others have suggested that lexical tone perception continues to develop beyond the preschool years (Wong, 2013; Wong & Strange, 2017). Given these findings, it is suggested that linguistic experience may have a significant role in shaping an individual's perceptions of speech sounds, in particular lexical tones.

Sometimes individuals, such as international adoptees, may have substantially different linguistic experiences across infancy and adulthood. For example, international adoptees from China adopted as infants into English speaking families in the United States move from a tonal to a non-tonal language environment. When these adoptees' early exposure to a tonal language is pitted against the observed decline in lexical tone discrimination amongst English-speaking infants (Mattock and Burnham, 2006), it is uncertain whether early-formed linguistic representations will remain accessible after contact with the heritage language is discontinued.

There has been little research into the long-term effects of discontinued language exposure in infancy, and the available results are equivocal as to whether such early-formed linguistic representations are maintained across development. For Korean international adoptees in France with no conscious recollection of Korean, there was no difference in brain activation when they listened to Korean sentences compared to sentences in unfamiliar languages (e.g., Polish, Japanese, and Wolof; Pallier et al., 2003). Furthermore, the adoptees perceived Korean phonetic differences no better than native French speakers with no Korean experience (Ventureyra et al., 2004). Likewise, a case study with a child adopted from China to Canada by a native English speaker demonstrated significant decline in Chinese comprehension and expression over time during Chinese play sessions and rapid acquisition of English during English play sessions (Nicoladis & Grabois, 2002). In contrast to these results, however, children adopted from China with no conscious recollection of Chinese recruited a similar neural pattern to that of Chinese/French bilingual speakers when they heard three-syllable sentences in Chinese (Pierce, et al., 2014), suggesting that early language experiences may have long-lasting effects on linguistic processing. As a result, it remains uncertain whether and to what extent early language experiences leave a trace on the auditory and speech perception systems after heritage language exposure is discontinued.

Differences in tone processing among native tonal and non-tonal language speaking adults are well established (e.g., Klein, Zatorre, Milner, & Zhao, 2001; Gandour, Wong, & Hutchins, 1998). Tonal languages, such as Chinese languages, Vietnamese, and Thai, use tonal differences (primarily instantiated via vocal pitch) to indicate word meaning in the same way that non-tonal languages use only consonants and vowels. Perception of lexical tones is influenced by linguistic experience, as demonstrated by the discrimination of Mandarin words with different tones by Mandarin speakers and not by non-tonal language speakers (Xu, Gandour, & Francis, 2006). When lexical tones are perceived as nonlinguistic, pitch processing is localized in the right hemisphere (Zatorre, Evans, Meyer, & Gjedde, 1992; Wong et al., 2004). In contrast, linguistically relevant processing is predominantly in the left-hemisphere, as demonstrated by increased activation in left-hemisphere regions associated with linguistic processing in Mandarin speakers when discriminating Mandarin tones compared to right-hemisphere activation in English speakers (Klein et al., 2001; Wong et al., 2004). Specifically, the left inferior frontal gyrus and left planum temporale (PT) have been implicated in lexical tone perception (Gandour et al., 1998; Hickock & Poeppel, 2007). Interestingly, there was significant overlap in the activation of the left PT among the international adoptees and Chinese speakers in the Pierce et al. (2014) study, suggesting that the international adoptees may have retained the ability to perceive lexical tones as linguistically relevant.

The previous literature on international adoptees relied on cortical measures; however, subcortical measures have been unexplored. The auditory brainstem response

(ABR) is a set of scalp-recorded auditory evoked potentials that can be measured from infants and adults (Jewett & Williston, 1971; Hecox & Galambos, 1974). Differences in auditory brainstem responses between tonal and non-tonal language speakers have been observed (e.g., Swaminathan, Krishnan, & Gandour, 2008). An auditory evoked response of particular interest in the neural encoding of lexical tones is the frequency following response (FFR), which is a scalp-recorded response that has been used to study the representation of pitch at the subcortical level (Worden & Marsh, 1968; Krishnan, Xu, Gandour, & Cariani, 2004). Native Mandarin speakers exhibited more accurate pitch tracking and greater pitch strength compared to native English speakers, suggesting that pitch representations at the subcortical level are also dependent on language experience (Swaminathan et al., 2008; Krishnan & Gandour, 2009). Research comparing FFR responses to voice pitch in Mandarin-speaking adults and Chinese infants has also shown that pitch representations mature early in human development (Jeng et al., 2010; Jeng et al., 2016). If pitch representations in the auditory brainstem mature in infancy, then Chinese international adoptees may retain higher-fidelity pitch representations into adulthood even after exposure to their heritage language has been discontinued.

Though characteristics of auditory evoked responses have been well studied, knowledge about the maturation and plasticity of the auditory system continues to be explored. Early research has suggested that development of subcortical responses progresses until two years of age, when maturation plateaus (Salamy, 1984). However, there have also been suggestions that the auditory brainstem continues to mature into adolescence (Johnson, Nicol, Zecker, & Kraus, 2008; Skoe, Krizman, Anderson & Kraus, 2015; Krizman et al., 2015). Cross-sectional studies have demonstrated that the auditory brainstem continues to develop into the preschool years from ages 3 to 5 years (Spitzer, White-Schwoch, Carr, Skoe, & Kraus, 2015), and that auditory brainstem maturation continues between ages 3 to 12 years (Skoe et al., 2015; Krizman et al., 2015). As a result, the plasticity of the subcortical auditory system can be evaluated with international adoptees who are exposed to their heritage language early in life and adopted as infants or young children. With no prior research on subcortical and behavioral differences in pitch processing by international adoptees from a tone language background, the impact of early language exposure on subcortical phonetic representations and behavioral differences between international adoptees and speakers of their heritage language remains unknown.

The present study tested adult international adoptees, who were adopted from China as infants or toddlers and raised in American English-speaking homes, to explore the plasticity of the subcortical auditory system in relation to lexical tone perception. International adoptees were compared to native English and native Chinese speakers. Listeners were presented with Mandarin monosyllables to elicit the FFR. We hypothesized that early tonal language experiences leave enhanced, unconscious subcortical pitch representations and that the FFR recordings for the international adoptees are similar to those of the native Chinese speakers. An alternative hypothesis is that the FFR of the international adoptees resembles these from the native English speakers, and that neurophysiological traces of early tonal language experiences are not retained at the subcortical level.

METHODS

Participants

Three groups of adult participants were recruited for the study: native American English (AE) adults (n = 21), native Mandarin Chinese (MC) adults (n = 21), and international adoptee (IA) adults, who, as children, were adopted from China to the United States (n = 9). The AE group was comprised of native American English speakers with no tonal language (e.g., Mandarin and other Chinese dialects, Thai, Vietnamese, etc.) experience. The mean age of the AE group was 21.2 years and was comprised of 17 female and 4 male participants. AE participants had an average of 4.9 years of musical experience. The MC group were native Mandarin speakers, who lived in Mainland China through the age of 18 and who lived in the United States for less than 4 years. The mean age of the MC group was 22.2 years and was comprised of 21 female and 1 male participant. MC participants had an average of 7.7 years of musical experience. The IA group consisted of all female individuals adopted from Mainland China by American English-speaking families in the United States. The IA group had mean age was 20 years and had an average of 9.9 years in musical experience.

Due to challenges with recruitment, international adoptees with minimal Mandarin exposure post-adoption were included in the study. Mandarin experience was quantified using the self-reported Mandarin experience and the Hanyu Shuiping Kaoshi (HSK) Level 1 test, which has listening and reading comprehension parts. Level 1 is designed for individuals who have some understanding and use some simple Chinese characters and sentences. Experience with Mandarin was considered during the analysis and is presented in **Table 1**.

To recruit participants, support group networks and adoption organizations in the United States were contacted via email. All participants were recruited from Boston University and the local community. All participants passed pure tone threshold testing; with inclusion criteria requiring hearing thresholds ≤ 20 dB HL at octaves frequencies from 250 to 8000. Participants with auditory, psychological, or neurological disorders were excluded from the study. Participants also completed an adapted version of the Edinburgh Handedness Inventory (Oldfield, 1971) and a language history questionnaire. Music experience often results in improved brainstem pitch tracking (Wong, Skoe, Russo, Dees, & Kraus, 2007); correspondingly, all participants completed a music history questionnaire. Participants with varying music backgrounds were included; the effect of music experience on behavioral measures was examined between the groups in post-hoc analyses.

Table 1. International Adoptee Demographic Information

Participant	Age (yr.)	Adoption Age (mo.)	Musical Experience (yr.)	Mandarin Experience	HSK score
p0846	19	6	8	None after adoption	N/A
p0998	21	3	14	None after adoption	N/A
p1040	18	10	14	1 year in grade school	28/40
p2207	21	6	7	6 years from high school to college	32/40
p4343	19	12	10	3 years in grade school	23/40
p5699	20	12	12	11 years since 3rd grade	39/40
p6870	20	18	8	Two years in college	39/40
p9031	23	8	12	2 years in middle school	14/40
p9945	21	11	7	None after adoption	N/A
Average	**20**	**9.6**	**9.9**		

Stimuli

A set of four minimally contrastive Mandarin monosyllables (/yi^1/ “clothing”, /yi^2/ “aunt”, /yi^3/ “chair”, /yi^4/ “easy”) was used to elicit the FFR. The monosyllables were minimally distinguishable by tone where tone 1 (T1) had a level pitch contour, tone 2 (T2) a rising pitch contour, tone 3 (T3) a bidirectional falling-rising pitch contour, and tone 4 (T4) a falling pitch contour. The stimuli were recorded by a male native Mandarin Chinese speaker at 44.1 kHz in a sound-attenuated chamber using a Shure MX153 earset microphone, Behringer MIC2200 microphone preamplifier, and Roland Quad Capture

USB sound card. The stimuli were normalized for intensity (70 dB SPL RMS amplitude) and duration (270 ms) using Praat. Fundamental frequency (f_0) contour was the acoustic feature principally distinguishing the monosyllables. A mean inter-stimulus interval of 300±100 ms was jittered from trial to trial so that signals unrelated to the acoustic stimuli differed at the start of each trial. Native Mandarin speakers listened to the stimuli to confirm that the stimuli were natural exemplars of the four monosyllables.

Design

This experiment occurred over two sessions either on two separate days or one day with a break. In the first session, participants listened to Mandarin monosyllables presented in random order while their FFR was measured. In the second session, participants performed a battery of behavioral tasks.

Procedure

FFR Task

First, electrodes were placed on the scalp of the participant, who sat in an acoustically and electrically shielded booth. Participants were asked to refrain from extraneous body movements to minimize artifacts. Throughout data collection, participants watched a silent movie to encourage a quiet yet wakeful state. FFR were recorded in response to diotic stimulation. The order of stimuli was randomized across and within subjects. Acoustic stimuli were presented through insert earphones with foam tips (ER-1, Etymotic, Elk Grove Village, IL) at a constant amplitude of 70 dB SPL. All stimulus delivery was controlled by specialized sound-control hardware (System 3 real-time signal processing systems, including D/A conversion and amplification; Tucker

Davis Technologies, Gainesville, FL).

During all test conditions, FFR responses were recorded at a sampling rate of 4096 Hz using a BioSemi Active Two System (BioSemi, Amsterdam, Netherlands). FFRs were recorded using a 32-channel electrode montage arranged according to the 10-10 system over the entire scalp. Two surface electrodes placed on the earlobes served as reference channels. Additionally, four external electrodes were used to measure vertical eye movements (EOG), which can affect EEG data. Electrode offsets for EEG and EOG were maintained below 30 μV. Approximately 6400 trials were collected from each participant, with 1600 trials per stimulus (T1, T2, T3, & T4). Of the 1600 trials, 800 trials were presented with reversed polarity, allowing us to differentiate between FFR phase-locked to the stimulus envelope and FFR phase-locked to the spectral components of the stimulus (Bharadwaj & Shinn-Cunningham, 2014).

Data Pre-Processing

First, EEG inputs were high-pass filtered with a 70 Hz low-frequency cutoff to isolate subcortical responses while minimizing signal related to cortical activity. After filtering, responses were averaged with an epoching window. Trials with peaks greater than 35μV were rejected to remove muscle and other artifacts. The remaining trials were averaged for each subject, with at least 975 trials per tone per subject (mean = 1475 trials, SD =133 trials).

Data Analysis

Two quantitative measures from FFR were used to calculate the accuracy and magnitude of pitch processing: pitch tracking accuracy and pitch strength. For each subject, the FFR pitch tracking and pitch strength were derived using autocorrelation. A sliding window analysis was used in which 40 ms bins of FFR data were analyzed in the lag domain. The 40 ms sliding window was shifted in 3 ms steps, resulting in 80 overlapping windows. Pitch tracking accuracy refers to how well the FFR follows changes in pitch and was measured using stimulus-to-response correlations. For each window, the maximum peak autocorrelation value was calculated. After the reciprocal was taken for corresponding time lags, which represented an estimate of the f_0. The time lags associated with autocorrelation peaks from each frame were concatenated together to make a pitch contour. All data analyses were performed using MATLAB R2017b (The MathWorks, Inc., Natick, MA)

Pitch strength was calculated using root-mean-square (RMS) values of the pitch tracking points. First the pitch estimates for the participant and the corresponding stimulus were aligned in time. After the difference between the points for the pitch estimates were calculated, the differences were averaged and then squared to obtain an RMS error value in Hz. The RMS error represents how far the participant's pitch estimation deviated from the stimulus pitch contour.

Statistical Analyses

Pitch Tracking. For pitch tracking accuracy, the stimulus-to-response association was calculated using a Pearson correlation coefficient (r) between the f_0 contours of the stimulus and FFR response. The correlation measures the strength and direction of the linear relationship between the stimulus and response f_0 contours. After the correlations were transformed from an r-value to z-value using the Fisher-transform, t-tests were used to determine if there were pairwise differences in pitch tracking between the groups for each tone.

Pitch Strength. Statistical differences in pitch strength correlations were determined using three linear mixed effects models. The dependent measure was the RMS error value that measured how well the FFR contour follows the f_0 contours of the stimulus. T-tests were used to determine if there were pairwise differences in pitch strength between the groups for each tone. All statistical analyses were performed using R (RStudio Team, Boston, MA).

Behavioral Tasks

Pitch-Contour Perception Test (PCPT). PCPT is a non-lexical task that is designed to assess the participants' pitch-perception abilities in relation to their ability to learn lexical tones. An accuracy of greater than 70% has been established as indicating a high aptitude for learning lexical tones (Wong & Perrachione, 2007). The participants listened to vowels created from recordings from four native American English speakers (two male, two female), five vowels (/a/, /i/, /o/, /e/, and /y/), three tones, and two repetitions in a sound-attenuated chamber with a Shure WH20-XLR microphone via a Roland UA25EX sound card sampling at 44.1 kHz with a 16-bit sampling depth. Using Praat, the stimuli were cut and normalized to 70 dB SPL RMS. Vowels were superimposed with a level, rising, or falling pitch contours using the pitch synchronous overlap-and-add algorithm (PSOLA) (Moulines & Charpentier, 1990) in Praat (Boersma & Weenik, 2018).

During the task, participants sat in an acoustically and electrically shielded booth, while they listened to the auditory stimuli. All participants were familiarized with the task before beginning the experiment. The participants completed 120 trials (4 talkers x 5 vowels x pitch contours x 2 repetitions). When they heard the auditory stimuli, participants saw different pitch contours on the screen (→ = level, ↗ = rising, and ↘ = falling) and were asked to use a keypad to match the pitch contour with the auditory stimulus they heard (e.g., press 1 =↗ (on left side of the screen) or press 2 = ↘ (on the right side of the screen)). Participants were not given feedback as to whether their response was correct (Perrachione, 2014).

Tone Discrimination Task. A task was conducted with all participants to assess their lexical tone discrimination using an AX paradigm. During the task, participants were presented with pairs of stimuli that differed in consonants, vowels, and speaker and were asked to decide whether the second sound ‘X’ had the same (e.g., mǎi / yě) or different (e.g. tīng / xià) pitch contour than the first sound ‘A’. Two conditions were used: Mandarin monosyllables (speech) and synthesized tones (sinewave) that removed the semantic effect of the Mandarin monosyllables. The stimuli were recorded by one female and one male native Mandarin speaker. The sinewave condition stimuli were synthesized using pitch contours from the natural speech tokens using the PSOLA in Praat. While sitting in an acoustically shielded booth, participants listened to 96 different Mandarin monosyllables. The stimuli were separated by an ISI of 0.5s. In total there were 192 trials, half with the same tone (12 pairs x 4 lexical tones x 2 conditions) and half with different tones (8 pairs x 6 different tone combinations x 2 conditions).

Stroop Task. Participants completed an auditory Stroop task to assess familiarity with Mandarin and Cantonese. The Stroop effect is a phenomenon that results in slower response times when there are incongruent stimuli presented simultaneously. For example, response inhibition occurs because of the strong association between the letters and the semantic meaning of the letters (i.e., the letters “r”, “e”, “d” are associated with the color “red”), which interfere with the process of having to choose the actual color of the letters. (Stroop, 1935; Dyer, 1973). Because of past auditory Stroop tasks demonstrating the effects of incongruent stimuli on response time (Spapé & Hommel, 2008), it is expected that participants familiar with the words will have greater interference, resulting in slower response time in incongruent trials than congruent trials for participants who are speakers of the language. This task will be used to assess whether the IA participants have latent Mandarin or Cantonese knowledge.

During the Stroop task, participants listened to the words “high” and “low” spoken in English, Mandarin, and Cantonese in a high-pitched and low-pitched voice by one male native speaker of American English, Mandarin, and Cantonese, respectively. The stimuli were recorded at 44.1 kHz in a sound-attenuated chamber using a Shure MX153 earset microphone, Behringer MIC2200 microphone preamplifier, and Roland Quad Capture USB sound card. The participants listened to 40 trials (1 speakers x 40 incongruent/congruent trials (20 congruent/20 incongruent) per language condition. Their task was to decide whether the word was said with a high-pitched or low-pitched voice. Participants were given feedback (i.e., a check or X) if their answer was correct or incorrect after each trial.

Sentence Identification Task. The sentence identification task was designed to assess if the adoptees retained remnants of their native language at different linguistic levels, from phonology to the lexicon, prosody, syntax and semantic level. Because the MC participants are native speakers of Mandarin, they will be used as the standard for indicating if an individual has access to the linguistic information necessary for recognizing Chinese sentences. It was hypothesized that the AE participants would be worse at identifying the Chinese sentences because of their unfamiliarity with Mandarin and Cantonese. Both Mandarin and Cantonese were chosen based on the possibility that adoptees may also have been adopted from Hong Kong or Macau, whose official languages are Cantonese, Mandarin, and English and Cantonese and Portuguese, respectively (Civil Service Bureau, 2017; “Geography and Population”, n.d.). If the adoptees perform similarly to the MC participants, this would be an indication that the participants have maintained traces of linguistic information of varying levels from their heritage language and have good identification of Chinese. If the adoptees perform poorly or more similarly to AE participants, then this would indicate poor maintenance of their heritage language.

All participants listened to sentences in Mandarin, Cantonese, Japanese, Korean, and Arabic to assess their ability to perceive Chinese. Sentences (Mandarin and Cantonese) from “The North Wind and Sun” passage were recorded by two female and two male native speakers of each language at 44.1 kHz in a sound-attenuated chamber using a Shure MX153 earset microphone, Behringer MIC2200 microphone preamplifier, and Roland Quad Capture USB sound card. The text for all the languages were obtained from the Handbook of the International Phonetic Association (International Phonetic Association, 1999). Participants listened to 4 phrases spoken by 4 different speakers in 5 languages, totaling 80 targets that were presented in random order across participants and at 70dB. Using a keypad, the participants decided whether the sentence was in Chinese by selecting "Chinese" or "not Chinese" on the screen.

Statistical Analysis

PCPT. The dependent variable on this task was accuracy (number of trials where the pitch contour was correctly selected out of the total number of trials). To determine pairwise differences between groups in pitch contour identification accuracy, three linear mixed effects models for binomial data were conducted, comparing the AE to MC group, AE to IA group, and the IA to MC group. Linear mixed effects models contained a fixed factor for group and random intercepts by participant. Significance of effects were determined by estimating degrees of freedom using the Satterthwaite method implemented in the package 'lmerTest' in R. We adopted a significance criterion of $\alpha = 0.05$.

Tone Discrimination. The dependent measure on this task was accuracy (number of trials where the participant made the correct choice). To determine pairwise differences between the groups in tone discrimination accuracy, three linear mixed effects models for binomial data were conducted, comparing AE to MC group, AE to IA group, and the IA to MC group. Linear mixed effects models contained a fixed factor for group and stimuli condition (Chinese, sinewave) and random intercepts by participant. Significance of effects were determined by estimating degrees of freedom using the

Satterthwaite method implemented in the package ‘lmerTest’ in R. We adopted a significance criterion of $\alpha = 0.05$.

Stroop Task. Three linear mixed effects models were conducted to determine pairwise differences between the groups in response time, comparing AE to MC group, AE to IA group, and the IA to MC group. The dependent measure of this task was response time (for each trial). Linear mixed effects models contained a fixed factor for Group, language condition (Mandarin, English, Cantonese), and mismatch condition with random intercepts by participant. Statistical significance was determined with a significance criterion of $\alpha = 0.05$, with p-values based on the Satterthwaite approximation of the degrees of freedom using the package ‘lmerTest’ in R.

Sentence Identification. Six linear mixed effects models were conducted to determine pairwise differences between the groups in probability of classifying recordings from the various languages as “Chinese”. Three models were used to compare the probability of identifying Mandarin as Chinese between the AE to MC group, AE to IA group, and the IA to MC group. Three models were used to compare the probability of identifying Korean as Chinese between the AE to MC group, AE to IA group, and the IA to MC group. The dependent measure was probability of classification as Chinese. The linear mixed effects models contained fixed effects for group and sentence language, with random intercepts by participant. Statistical significance was determined with a significance criterion of $\alpha = 0.05$, with p-values based on the Satterthwaite approximation of the degrees of freedom using the package ‘lmerTest’ in R.

Effect of Demographic Characteristics. The effect of musical experience, age of adoption, HSK score, and amount of Mandarin re-exposure (i.e., the number of years of Mandarin experience an adoptee has after adoption) on PCPT score, tone discrimination accuracy, the Stroop Effect, and Sentence Identification accuracy were examined using a Pearson correlation. Additionally, the associations between musical experience, age of adoption, HSK score, and amount of Mandarin re-exposure were explored using a Pearson correlation in R.

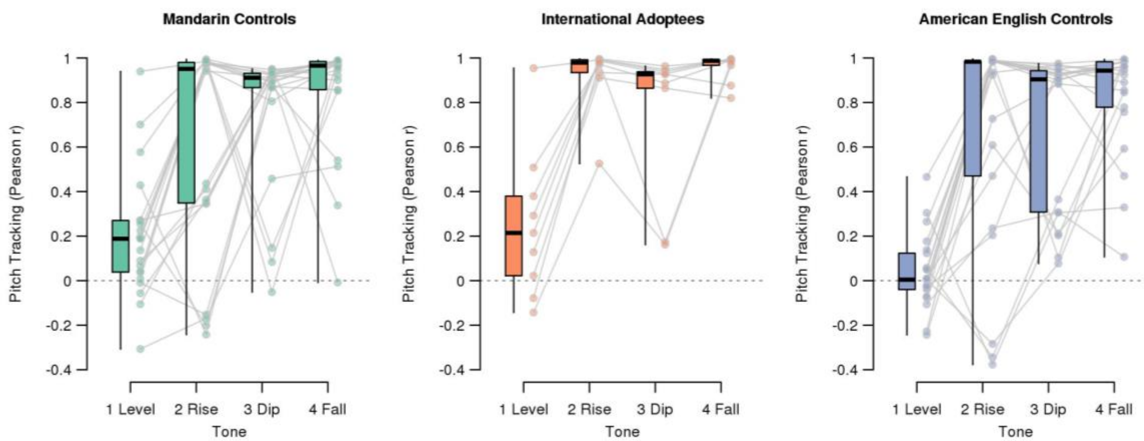
RESULTS

FFR Task

All three groups had similar tracking patterns (see **Figure 1**) and pitch strength (see **Figure 2**). The average stimulus-to-response correlation for the AE group was 0.03 ± 0.17 for tone 1, 0.67 ± 0.49 for tone 2, 0.70 ± 0.35 for tone 3, and 0.82 ± 0.25 for tone 4. For the IA group, the average stimulus-to-response correlation was 0.25 ± 0.34 for tone 1, 0.92 ± 0.15 for tone 2, 0.75 ± 0.33 for tone 3 and 0.96 ± 0.06 for tone 4. For the MC group the average stimulus to response correlation was 0.29 ± 0.28 for tone 1, 0.61 ± 0.47 for tone 2, 0.77 ± 0.31 for tone 3, and 0.84 ± 0.27 for tone 4. T-tests were used to determine if there were pairwise differences in pitch tracking and pitch strength amongst the groups for each tone. Compared to the AE group, the MC group had significantly different tone 1 tracking ($t = 2.08$, $p = 0.044$) and similar tone 2 ($t = -0.65$, $p = 0.052$), tone 3 ($t = 0.36$, $p = 0.72$), and tone 4 ($t = 0.35$, $p = 0.73$) tracking. The MC and IA group had similar tone 1 tracking ($t = -0.60$, $p = 0.56$), different tone 2 ($t = -1.91$, $p = 0.066$) tracking , similar tone 3 ($t = -0.11$, $p = 0.91$) tracking, and different tone 4 ($t = -1.91$, $p =$

0.066) tracking. The IA and AE group had significantly different tone 1 tracking (t = 2.27, p = 0.031), similar tone 2 (t = 1.27, p = 0.21) tracking, similar tone 3 (t = 0.35, p = 0.73) tracking, and significantly different tone 4 (t = 2.12, p = 0.042) tracking.

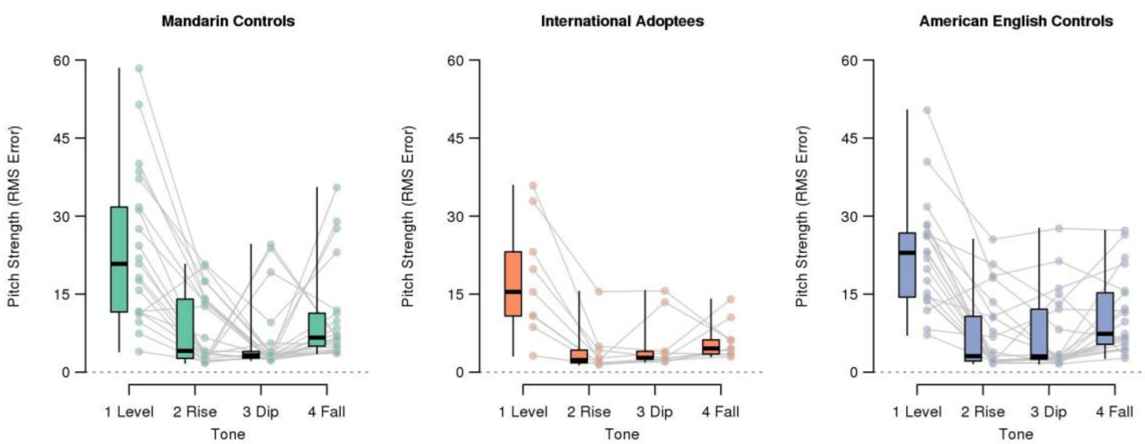
Figure 1. Stimulus-to-response correlations for each participant group (Mandarin, international adoptees, and American English) for each tone (1/level, 2/rising, 3/dipping, 4/falling.



The average RMS value of error for the AE group was 22.6 Hz ± 10.4 Hz for tone 1, 7.63 Hz ± 7.45 Hz for tone 2, 7.12 Hz ± 7.48 Hz for tone 3, and 11.0 Hz ± 7.59 Hz for tone 4. For the IA group, the average RMS error was 17.8 Hz ± 11.1 Hz for tone 1, 4.04 Hz ± 2.47 Hz for tone 2, 5.44 Hz ± 5.23 Hz for tone 3, and 6.17 Hz ± 3.74 for tone 4. For the MC group the average RMS error was 23. 9 Hz ± 14.8 Hz for tone 1, 8.60 Hz ± 6.90 Hz for tone 2, 6.23 Hz ± 7.06 Hz for tone 3, and 10.8 Hz ± 9.42 Hz for tone 4. Regarding pitch strength, the MC and AE groups had similar pitch strength for tone 1 (t = 0.35, p = 0.75), tone 2 (t = 0.44, p = 0.67), tone 3 (t = -0.40, p = 0.68), and tone 4 (t = -0.055, p = 0.96). The MC and IA groups also had similar pitch strength for tone 1 (t = 1.10, p = 0.28), tone 2 (t = 1.82, p = 0.080), tone 3 (t = 0.30, p = 0.77), and tone 4 (t = 1.42, p =

0.17). The IA and AE groups also had similar tone 1(t = -1.12, p = 0.27), tone 2 (t = -1.34, p = 0.19), tone 3 (t = -0.61, p = 0.55), and tone 4 (t = -1.79, p = 0.085). Based on the results, the MC group had significantly better tone 1 pitch tracking than the AE group, the IA group had significantly better tone 1 and 4 pitch tracking than the AE group, and the IA group had better tone 2 and 4 tracking than the MC group. Pitch strength was similar between the groups for each tone.

Figure 2. Root-Mean-Square error value for all participant groups (Mandarin, international adoptees, and American English) for each tone (1/level, 2/rising, 3/dipping, 4/falling).



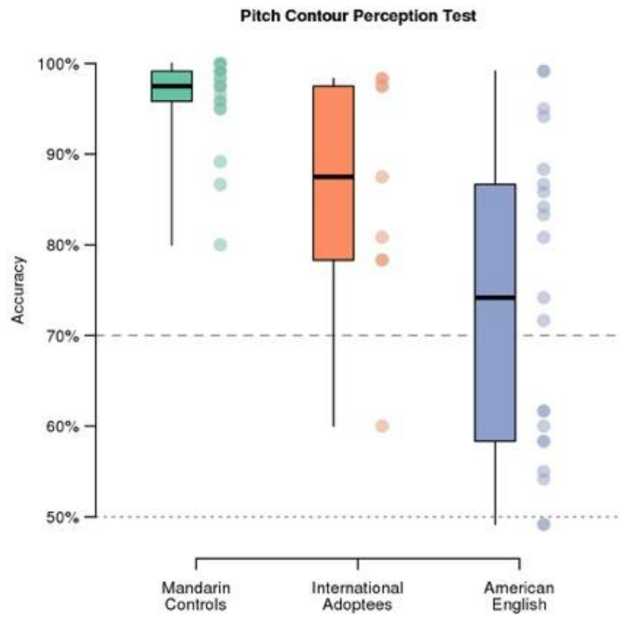
Behavioral Tasks

PCPT. Performance on the PCPT task showed variation in pitch perception abilities across the three participant groups (see **Figure 3**). Average performance for the MC group was 96.2% ± 0.049%, 86.3% ± 0.12% for the IA group, and 73.8% ± 0.17% for the AE group. Linear mixed effects models were conducted to determine differences in accuracy by group. The first model determined that the MC group performed significantly better than both the IA group (t = -3.01, p << 0.001) and the AE group (t = -

5.75, $p << 0.001$). The accuracy of the AE and IA groups were not significantly different ($t = -1.95$, $p = 0.06$), but there was a trend towards higher PCPT scores in the IA group. A Pearson correlation was conducted to examine the relationship between PCPT score and musical experience, age of adoption, Mandarin re-exposure (i.e., how much Mandarin the adoptee was exposed to after adoption), and HSK score in the IA group. For the IA group, there was no significant correlation between PCPT score and age of adoption ($r = 0.23$, $p = 0.55$), Mandarin re-exposure ($r = 0.05$, p = 0.90), or HSK score ($r = 0.02$, $p = 0.94$). Though not significant, there was a trend toward higher PCPT scores with more musical experience musical experience ($r = 0.54$, $p = 0.11$) in the IA group. Similarly, the musical experience had a significant positive effect on PCPT score in the AE group ($r = 0.60$, $p = 0.0041$).

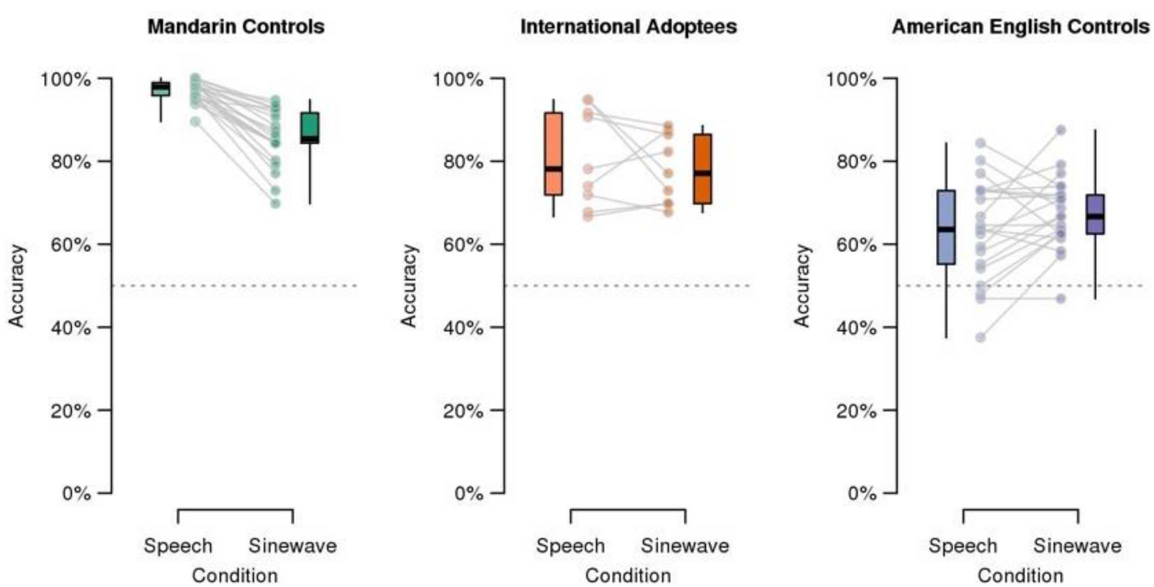
Based on the results, the MC group demonstrated significantly better pitch perception compared to both the IA and AE groups. The IA group also demonstrated better pitch perception abilities compared to the AE group. Musical experience had a positive effect on pitch perception abilities in the IA group.

Figure 3. Accuracy of each group (English, international adoptees, and Mandarin) on the pitch contour perception task.



Tone Discrimination. All three groups demonstrated varying abilities at tone discrimination (see **Figure 4**). The MC group scored an average of 97.2% ± 0.025% on the speech condition (i.e., Chinese words spoken by a native Mandarin speaker) and 85.7% ± 0.067% on the sinewave (i.e., synthesized tones) condition. The IA group scored an average of 81.8% ± 0.12% on the Chinese condition and 78.5% ± 0.081% on the sinewave condition. The AE group score an average of 63.8% ± 0.12% on the Chinese condition and 67.6% ± 0.083% on the sinewave condition. A linear mixed effects model was used to determine difference in accuracy based on condition and by group interactions. The first model compared the AE to the MC participants and determined that the MC participants were significantly better at tone discrimination than the AE participants (t = -14.06, p << 0.001) and the presence of condition and group interactions (t = 8.98, p <<0.001), such that Mandarin listeners showed better performance in Speech than Sinewave tones, whereas the AE group tended to show the reverse pattern. The second model comparing the IA and MC groups determined that the MC group was significantly better than the IA group (*t* = -6.31, *p* << 0.001) and that there were group and condition interactions. The third model comparing the AE and IA groups determined that the AE group was significantly better than the AE group (*t* = -4.60, *p* << 0.001) and that there was a group and condition interaction (*t* = 2.78, *p* = 0.005), such that the AE group was significantly better in the sinewave condition than the speech condition (*t* = 2.73, *p* = 0.006), but the IA group had no significant differences based on condition (*t* = -1.64, *p* = 0.10). Linear mixed effects models investigating the main effect of the conditions on the group determined that the MC group was significantly better in the speech condition than in the sinewave condition (*t* = -13.10, *p* << 0.001).

Figure 4. Tone discrimination accuracy by group (Mandarin controls, English controls, and International Adoptees) and condition (Speech and sinewave).



A Pearson correlation was conducted to examine the relationship between tone discrimination accuracy and musical experience, age of adoption, Mandarin re-exposure, and HSK score in the IA group. There was a strong positive correlation between speech tone discrimination and HSK score (r = 0.78, p = 0.01), but no significant correlation between sinewave tone discrimination and HSK score (r = 0.14, p = 0.72). There was no significant correlation between amount of Mandarin re-exposure and either speech (r = 0.44, p = 0.23) or sinewave (r = -0.091, p = 0.82) tone discrimination. Additionally, there was no significant correlation between age of adoption and speech (r = 0.63, p = 0.068) or sinewave (r = 0.31, p = 0.42) tone discrimination; although the relationship between adoption age and speech tone discrimination was strongly positive and in the expected direction. Furthermore, there was no significant correlation between music experience and both speech (r = 0.46, p = 0.21) and sinewave (r = 0.39, p = 0.30) tone discrimination in the IA group. Similarly, there was no significant correlation between music experience and both speech (r = 0.33, p = 0.14) and sinewave (r = 0.35, p = 0.68) tone discrimination in the AE group.

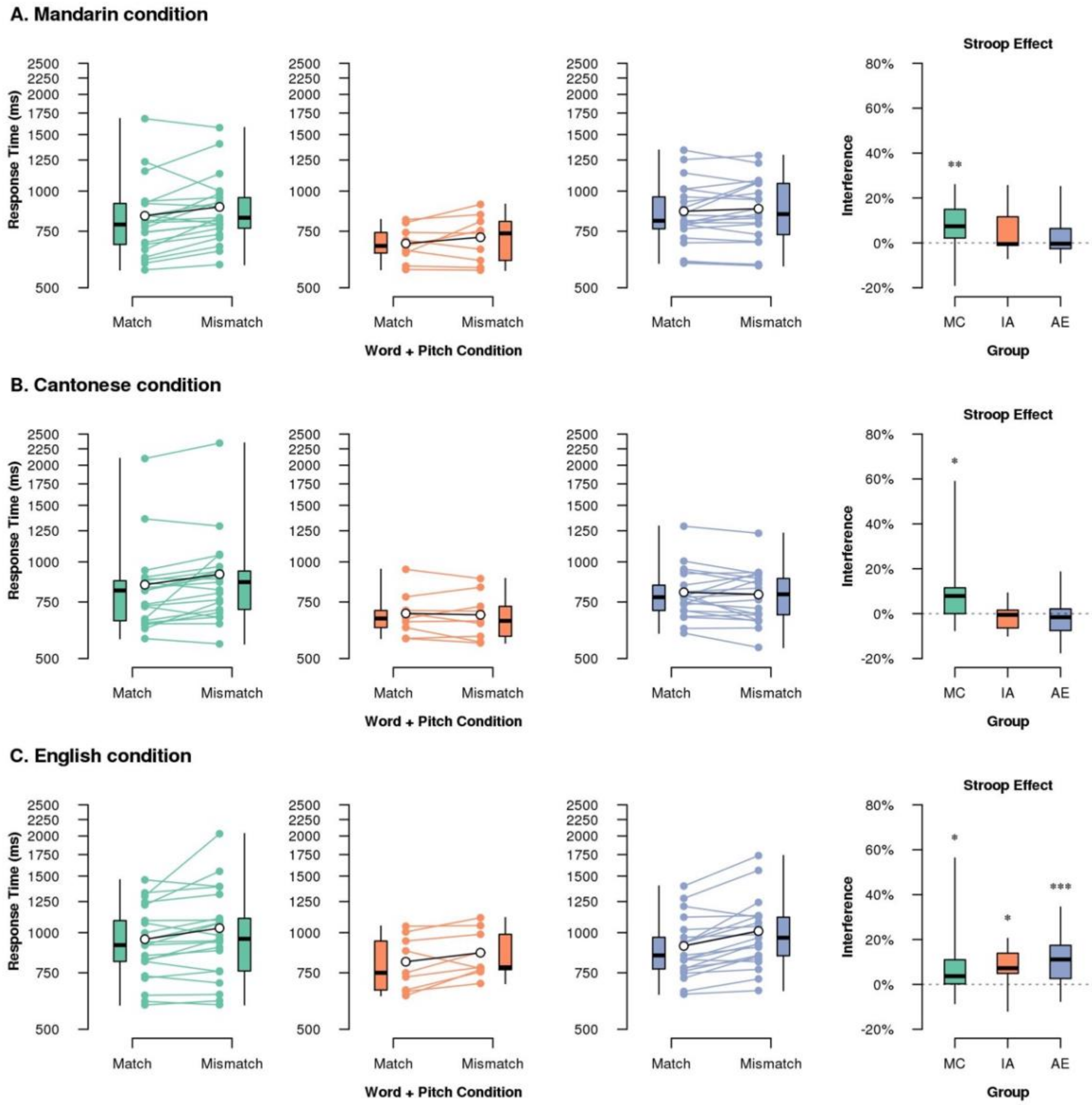
Together the results show that MC group was significantly better at tone discrimination than both the AE and IA groups; they were also significantly better at the speech condition than the sinewave condition. Furthermore, the IA group was significantly better at tone discrimination than the AE group, but performed similarly on the speech and sinewave conditions. Additionally, the AE group was significantly worse at tone discrimination than both the MC and IA groups and performed significantly better with sinewaves. In the IA group, a higher HSK score was associated with better speech tone discrimination. However, HSK score was not correlated with sinewave tone discrimination. In addition, amount of Mandarin re-exposure, age of adoption, and musical experience were not significantly correlated with tone discrimination. However, there was a positive trend in the discrimination of speech tones and age of adoption.

Stroop Task. All participant groups demonstrated varying interference effects when there was a mismatch (i.e., “high” said with a low-pitched voice) vs. match (i.e., “high” said with a high-pitched voice) in stimuli, depending on the language condition (see **Figure 5**). The MC group had an average interference (i.e., difference in reaction time between mismatched and matched stimuli) of 55.0 ms ± 110.0 ms in the Mandarin condition, 66.6 ms ± 103.4 ms in the Cantonese condition, and 78.8 ms ± 176.4 ms in the English condition. The IA group had an average interference of 31.2 ms ± 72.1 ms in the Mandarin condition, -7.0 ms ± 45.5 ms in the Cantonese condition, and 54.3 ms ± 72.6 ms in the English condition. Finally, the AE group had an average interference of 15.8 ms ± 72.6 ms in the Mandarin condition, -12.7 ms ± 69.8 ms on the Cantonese condition, and 102.4 ms ± 79.3 ms in the English condition.

Linear mixed effects models were conducted to determine whether groups demonstrated a significant difference in reaction time between “mismatch” and “match” stimuli for each condition. In the Mandarin condition, the MC group demonstrated a significant difference ($t = 3.36$, $p << 0.001$), the AE group demonstrated no significant difference ($t = 1.20$, $p = 0.23$), and the IA group also demonstrated no significant difference ($t = 0.82$, $p = 0.42$) between match and mismatch stimuli reaction times. Similarly, in the Cantonese condition, the MC group demonstrated a significant difference ($t = 3.30$, $p << 0.001$), the AE group demonstrated no significant difference ($t = -0.27$, $p = 0.79$), and the IA group also demonstrated no significant difference ($t = -0.76$, $p = 0.45$) between match and mismatch stimuli reaction times. In the English condition, all groups demonstrated a significant difference: MC group ($t = 2.42$, $p = 0.02$), IA group ($t = 2.56$, $p = 0.01$), and AE group ($t = 5.13$, $p << 0.01$) between match and mismatch stimuli reaction times

Pearson correlations were conducted to determine the relationship between Stroop effect in the Mandarin condition to HSK score, amount of Mandarin re-exposure, and age of adoption in the IA group. There was no significant correlation between reaction time differences in the Mandarin condition and HSK score ($r = -0.14$, $p = 0.72$), amount of Mandarin re-exposure ($r = 0.20$, $p = 0.17$), and age of adoption ($r = -0.50$, $p = 0.17$). Based on the results, only the MC group demonstrated a Stroop effect in the Mandarin and Cantonese conditions. All groups demonstrated a Stroop effect for the English condition, with the AE and IA groups tending to show a larger effect than the MC group. Finally, reaction time differences in the Mandarin condition for IA participants was unrelated to HSK score, amount of Mandarin re-exposure, and age of adoption.

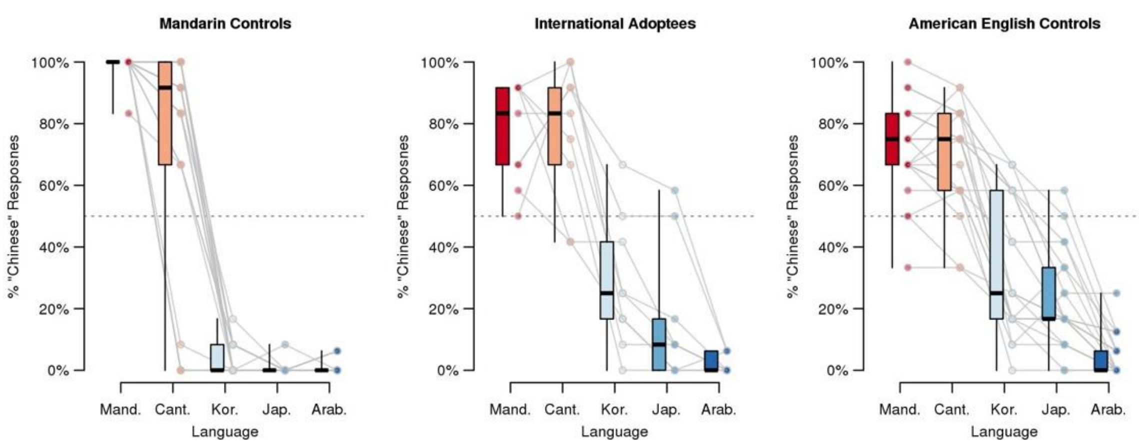
Figure 5. Reaction times for word and pitch condition (match and mismatch) by group (Mandarin (MC), international adoptees(IA), and American English (AE) and language condition (Mandarin, Cantonese, and English) and the Stroop effect



Sentence Identification. The MC group demonstrated greater accuracy at identifying Chinese sentences with an average hit-rate (i.e., selected phrases said in Mandarin or Cantonese) of 85.8% % ± 17.6% and an average false-alarm rate (i.e., selecting Korean, Japanese, or Arabic phrases as Chinese) of 1.4% ± 1.85%. The IA group had an average hit-rate of 77.4% ± 11.7% and a false alarm rate of 13.9% ± 12.8%. The AE group demonstrated the lowest accuracy at identifying Chinese sentences with an average hit-rate of 70.1% ± 14.9% and a false-alarm rate of 18.6% ± 11.9%.

Linear mixed effects models were conducted to determine if there were significant differences between selecting Mandarin or Korean sentences as Chinese. Korean was chosen for this model because it was the most frequent language mistaken as Chinese (see **Figure 6**). Amongst the MC and AE groups, the MC participants were significantly more likely to identify Mandarin phrases as Chinese (t = -8.00, $p \ll 0.001$) whereas AE participants were significantly more likely to identify Korean as Chinese (t = 6.21, $p \ll 0.001$). Amongst the IA and AE groups, both groups had similar tendencies to identify Mandarin phrases as Chinese (t = -0.86, p = 0.40) and Korean phrases as Chinese (t = 0.64, p = 0.53). Compared to the IA group, the MC group was more likely to identify Mandarin phrases as Chinese (t = -5.991, $p \ll 0.001$) whereas the IA group was more likely to identify Korean phrases as Chinese (t = 5.202, $p \ll 0.001$).

Figure 6. Percentage of targets identified as Chinese when hearing phrases in Mandarin, Cantonese, Korean, Japanese, and Arabic.



A Pearson correlation was conducted to determine the relationship between the probability of identifying Mandarin as Chinese and HSK score, years of Mandarin re-exposure, and age of adoption in the IA group. A Pearson correlation was also conducted to determine the relationship between misidentifying Korean as Chinese and HSK score, amount of Mandarin re-exposure, and age of adoption. The HSK score had a strong positive correlation with the correct identification of Mandarin phrases ($r = 0.83$, $p = 0.0052$) and a strong negative correlation with identifying Korean as Chinese ($r = -0.83$, $p = 0.0062$). The probability of identifying Mandarin sentences as Chinese was not significantly correlated with amount of Mandarin re-exposure ($r = 0.54$, $p = 0.13$) and age of adoption ($r = 0.43$, $p = 0.25$). Additionally, there was a moderate negative correlation between the amount of Mandarin re-exposure and misidentifying Korean ($r = -0.69$, p = 0.039). There was no correlation between age of adoption and misidentifying Korean ($r = -0.48$, $p = 0.19$).

Based on the results, the MC group was significantly better at identifying sentences in Chinese compared to the IA group and the AE groups, who were likely to mistake Korean as Chinese. A higher HSK score was associated with higher likelihood of correctly identifying Mandarin sentences. Furthermore, more Mandarin exposure after adoption and a higher HSK score was associated with less misidentification of Korean as Chinese.

Effects of Demographic Characteristics. Pearson correlations between the music experience, age of adoption, HSK score, and Mandarin re-exposure were examined to determine associations between the variables. There was a strong positive association between HSK score and years of Mandarin re-exposure ($r = 0.71$, $p = 0.034$). There was no significant association between HSK score and age of adoption ($r = 0.61$, $p = 0.078$); however, there was a positive trend between the two variables. There was no significant relationship between Mandarin re-exposure and ago of adoption ($r = 0.50$, $p = 0.26$). There was no significant association between musical experience and Mandarin re-exposure ($r = 0.029$, $p = 0.94$), HSK score ($r = 0.014$, $p = 0.97$), or age of adoption ($r = 0.26$, $p = 0.50$). Based on the results, performance on the HSK test was related to amount of Mandarin re-exposure. There was no relation between musical experience and all other variables.

DISCUSSION

This study examined whether individuals removed from their native language (i.e., international adoptees) at a young age maintained traces of their native language even after a prolonged period of minimal to no exposure to it, as measured by auditory FFR and behavioral tasks. Analysis of the FFR data demonstrated that all participant groups tracked the pitches similarly. Behaviorally, the IA participants had significantly poorer pitch perception compared to MC participants and similar pitch perception to AE participants. Additionally, the IA participants demonstrated poor tone discrimination compared to MC participants, but were significantly more accurate than AE participants. Furthermore, the IA group demonstrated a Stroop effect only for the English condition. Finally, the sentence identification task demonstrated the IA participants had poor Chinese sentence identification compared to MC participants and were similar to AE participants regardless of Mandarin experience after adoption.

To acquire a language, we must learn the phonetic distinctions that will be utilized by that language. From birth, infants can discriminate more sound contrasts than their native language (Eimas et al., 1971; Kuhl, 2004). Nearing twelve months, infants to demonstrate an increasingly more adult-like way of phonetic discrimination, which is the ability to discriminate your native language contrasts better than non-native language contrasts (Werker & Lalonde 1998). As the infants acquire their phonemic inventory, acquisition of the rhythm and intonation of a language are demonstrated in hierarchical progression of babble (Whalen, Levitt, & Wang, 1991). With international adoptees, their language environment is modified early in life. At first, the adoptees learn from language input heard in the womb and early language experience in their native country. Then, after adoption and in the early months of their life, their linguistic input changes to that of their adoption country and to the language they eventually acquire and use in their adult lives. Adoptees often acquire their adoptive language quickly and like infants in their adoptive country (Snedeker, Geren, & Shafto, 2007), and often master their adoptive language within one to two years (Roberts et al., 2005). However, with their adoptive language acquisition, their heritage language is rapidly lost, and children begin to forget words within months of adoption (Isurin, 2000; Nicoladis & Grabois, 2002). As adults, these adoptees with no native language re-exposure report no conscious recollection of their native language (Pallier et al., 2003; Ventureyra et al., 2004).

In this experiment, we explored the questions of whether early language learning is preserved after removal from or minimal re-exposure to their heritage language after being adopted into a country with a different language. We explored if the adoptees preserved their heritage language at the brainstem level through FFR, which has been demonstrated to mature early in life (Jeng et al., 2010; Jeng et al., 2016). However, a definite consensus on the maturation of the subcortical system has been debated as to whether the auditory system matures until two years of age or into adolescence (Salamy, 1984; Johnson et al., 2008; Skoe et al., 2013; Krizman et al., 2015). The preservation of differing levels of linguistic features from phonetics to phonology to semantic knowledge to suprasegmental characteristics of speech were explored with the behavioral tasks.

Through the FFR task, the IA participants demonstrated poor tone one pitch tracking, very strong tone two pitch tracking, strong tone three pitch tracking, and very strong tone four pitch tracking. Similarly, MC participants demonstrated poor tone one tracking and very strong tone two, three, and four pitch tracking. AE participants demonstrated similarly poor tone one pitch tracking and very strong tone two and four pitch tracking; however, the group had poor tone three tracking. The poor pitch tracking for tone one amongst the MC participants and contradictory behavioral performance is supported by Yu and Zhang, who found that FFR did not necessarily correlate with behavioral performance (2018). Additionally, poor tone one tracking is consistent with previous studies where level tone tracking was less robust than falling tone tracking (Krishnan et al. 2004, Jeng et al., 2010). Based on the results, it is difficult to come to a definite conclusion given the poor pitch tracking amongst the Mandarin control group; however, the Adoptee group showed consistently better pitch tracking and pitch strength than their native English-speaking peers.

The PCPT task was used to predict an individual's ability to learn a language with lexical tone contrasts via pitch identification performance (Wong & Perrachione, 2007). The IA group demonstrated poor pitch identification abilities compared to the MC participants, suggesting that their pitch processing capabilities are not as good as native Mandarin speakers. There was a positive trend towards higher PCPT score for the IA group compared to the AE speakers, suggesting that though adoptees have pitch perception abilities that are not the same as native Mandarin speaker, they are still better than non-tonal language speakers. Given that their HSK score, age of adoption, and amount of Mandarin re-exposure, had no significant effect on pitch perception abilities, this lack of correlation suggests that the adoptee's prior exposure to Mandarin was the primary source of their modest advantage over the AE participants with no tonal language exposure. Musical experience had an insignificant, but positive effect on pitch perception performance. Furthermore, the non-significant relationship between PCPT score and amount of Mandarin re-exposure suggests that re-exposure to Mandarin may not necessarily guarantee pitch perception abilities that are like a native tonal language speaker. However, there has been evidence indicating that adult non-tonal language speakers can learn to discriminate Mandarin lexical tones with extensive training (Reetzke, Xie, Llanos, & Chandrasekaran, 2018). Additionally, a score above 70% on the PCPT test suggests a high aptitude for learning lexical tones (Perrachione, Lee, Ha, & Wong, 2011). As a result, the PCPT scores aligns with previous research suggesting that adoptees have minimal sensitivity to the phonetics of their heritage language after being removed from their motherland (Pallier et al., 2003; Ventureyra et al., 2004).

Phonological processing of speech was explored with the tone discrimination task, which determined an individual's ability to discriminate pitches in speech and in conditions without linguistic information. This test was principally phonological, rather than phonetic, because it asked listeners to discriminate pitch contours across changes in syllable and talker, rather than in identical vowels and identical talkers, as in the PCPT. The MC group demonstrated significantly more accurate tone discrimination than the IA and AE groups with a significantly better accuracy in the speech condition. Interestingly, the IA group was significantly better at discriminating tones in both conditions than the AE group. Furthermore, their performance was unrelated to Mandarin re-exposure after adoption or age of adoption, suggesting that the adoptees are relying on knowledge unrelated to their recent Mandarin experiences. Additionally, tone discrimination was unrelated to musical experience in both the adoptee and English groups. However, performance in the IA group was associated with their HSK score, which was highly correlated with an adoptee's amount of Mandarin re-exposure, suggesting that their performance may have been influenced by the familiarity they obtained with Mandarin later in life. The MC group demonstrated a condition effect, suggesting that the MC group may be relying on linguistic information (i.e., semantics, stress patterns) beyond the phonetic information from the tone contrasts to make tone discrimination judgements. Contrastingly, the IA group demonstrated equal performance in both conditions, suggesting that their good tone discrimination abilities may be a result of good phonetic discrimination (i.e., pitch contours) that is superior to individuals without tonal language experience. The absence of a condition effect suggests that the IA group may not have access to higher level linguistic characteristics (i.e., semantics, stress patterns). Together these results support Ventureyra et al. (2003) and Pallier et al.'s (2004) studies because the adoptees do not demonstrate phonological processing to the level of native-speakers of Mandarin. The adoptees' superior performance to the AE group suggest that the adoptees may have minimal access to early acquired linguistic characteristics, such as phonology, if they are re-exposed to Mandarin.

The auditory Stroop task was based on the Spapé and Hommel's task (2008), which demonstrated that individuals reacted slower (i.e., demonstrated interference) when given an incongruent stimuli combination (i.e. low tone and "high" word). This behavioral task was used to determine if the IA participants would demonstrate a Stroop effect when presented with incongruent stimuli. It was hypothesized that the MC participants would demonstrate a Stroop effect during the English, Cantonese because the words "high" and "low" are linguistically similar (dai^1 and gou^1 vs. dī and gāo), and Mandarin condition. Additionally, it was hypothesized that the English participants would demonstrate a Stroop effect for only the English condition. The Stroop effect results from interference that results from competing lexical information. In this experiment, the IA and AE groups demonstrated a Stroop effect only in the English condition, whereas the MC group demonstrated a Stroop effect in all three conditions. The performance of the IA group on the Mandarin and Cantonese conditions suggests that their performance is unlike native Mandarin speakers whose interference results from extensive familiarity with the language, which results in deep connections between the phonetic characteristics of "high" and "low" and the semantic meanings associated with the words, which is weaker than the strong connections the adoptees have between the phonetic characteristics and semantic meanings of the words in their L2 (i.e., English). The loss of their heritage language's words has been demonstrated within months of adoption and into adulthood (Isurin, 2000; Nicoladis & Grabois, 2002; Pallier et al., 2003; Ventureyra et al., 2004), which likely result in weaker semantic connections between the word and the associated semantic meaning.

The sentence identification task was created to judge if the IA participants were better at identifying Chinese sentences from languages (Korean, Japanese, and Arabic) they were unfamiliar with, similar to Pallier et al.'s sentence identification task (2003). Additionally, this experiment was used to judge the adoptees' familiarity with Chinese given that some of the adoptees had post-adoption Mandarin exposure. The IA participants were no better at identifying the sentences than the AE participants and were significantly worse at identifying the Chinese sentences than the MC participants. Considering their experience with Mandarin after adoption, sentence identification affected by HSK score. In light of their overall inferior performance, the results support Pallier et al. (2003), who suggested that the adoptees have poor access to their heritage language at the phrasal level.

Overall, the findings of this study suggest that the IA participants have minimal access to higher-level representations of their heritage language even with some re-exposure to their heritage language. Because of the equivocal FFR results, conclusions about the maturation of the auditory brainstem are difficult to make. However, given the overall good FFR tracking results of the IA group, as well as their moderately better performance on the PCPT than the AE group, the results suggest that the participants may have access to pitch contour processing at the auditory brainstem. Behaviorally, the results support Pallier et al. (2003) and Ventureyra et al. (2004), who both suggest that individuals who were removed from their heritage language for a prolonged period with minimal exposure have minimal access to their heritage language. Additionally, the analysis of adoptees' linguistic experiences with Mandarin (i.e., HSK score, amount of Mandarin re-exposure, and age of adoption), suggest that the amount of Mandarin re-exposure may aid with phonological processing. Considering, one adoptee had eleven years of Mandarin experience post-adoption, the results suggest that Mandarin re-exposure does not guarantee native-like acquisition of their heritage language regardless of the amount of re-exposure. Considering that the average age of adoption of the IA participants who had minimal re-exposure to Mandarin was 7.5 months, their performance is consistent with Kuhl et al. (2001), who suggest a critical period from 6 to 12 months where infants are more attuned to the language in their environments. Though infants begin with the general ability to discriminate all the sounds in human languages (Eimas et al., 1971; Kuhl, 2004). These results support Kuhl's argument that the "critical period" results in the "neural commitment" to linguistic characteristics that are specific to their linguistic environment (Kuhl et al., 1992; Werker & Tees, 1984). Because of "neural commitment", the adoptees heritage language is mostly replaced by their adoptive language since their heritage language does not conform to the linguistic characteristics of the individual's adoptive language environment (Kuhl, et al., 1992; Werker & Tees, 1984).

The current study also provides evidence that when working with international adoptees who have exposure to their heritage language after adoption, the use of standardized testing is important for measuring an individual's true language abilities. In this study, the HSK score was highly correlated to years of Mandarin experience after adoption. However, analysis of tasks, such as the sentence identification and tone discrimination tasks suggested that the HSK score in comparison was a better measure of the adoptee's true Mandarin abilities.

Considering these conclusions, the FFR, PCPT, and tone discrimination results suggest that the adoptees may have an advantage when relearning Mandarin. This hypothesis is supported by their performance on the PCPT task, which is correlated with lexical tone language learning success (Wong & Perrachione, 2007), and their performance on the tone discrimination task. Additionally, Carcagno and Plack (2011) have found that increased FFR strength from training English participants with an auditory discrimination task was related to improved tone discrimination. Furthermore, studies on adult Korean international adoptees in Sweden, who had not been exposed to Korean after adoption and prior to the study, demonstrated more robust re-learning of Korean compared to their Swedish counterparts, who had more Korean training (Hyltenstam, Bylund, Abrahamsson, & Park, 2009). However, these results are contradicted by Ventureyra et al. (2004), who found that Korean adoptees with Korean re-exposure had no advantage in Korean phoneme discrimination. Furthermore, a study on Chinese adoptees in the Netherlands reported that the adoptees forgot the phonology of their heritage language shortly after adoption and did not have an advantage in tone discrimination after re-exposure (Zhou and Boersma, 2014). As a result, this is one aspect that can be investigated further given the equivocal results in this study and in others.

LIMITATIONS

The current study had a small sample size for all participant groups, especially the IA group. As a result, strong conclusions cannot be drawn on these data given the low power of this sample size, and replications are called for. Though one may argue that the inclusion of adoptees with Mandarin exposure after adoption limits the conclusions of this study, correlations between their experience after adoption and their behavioral tasks suggest that their Mandarin experience had little effect on their performance on the behavioral tasks, and when it did so it was specific to high-level Mandarin processing (e.g., Mandarin sentence identification).

The uncertainty of the FFR data limit our ability to draw conclusions on the maturation of the auditory brainstem. Poor tone one tracking is supported by the less robust tone one tracking seen in previous studies (Krishnan et al., 2004; Jeng et al., 2010). However, the FFR provides details on an individual’s ability to track pitches and may, therefore, provide insight into an individual’s ability to learn languages with lexical tone contrasts.

CONCLUSIONS

In conclusion, this study examined the FFR encoding and perceptual abilities of Chinese international adoptees with minimal to moderate re-exposure to Mandarin after adoption. Participants listened to Mandarin monosyllables representing the four tones, while the FFR was collected. Then participants completed behavioral tasks to determine pitch perception and tone discrimination abilities and their ability to identify Chinese languages amongst foreign languages. The FFR results indicated mostly good pitch tracking abilities that may support future tonal language learning in the adoptees. The behavioral data suggest that the adoptees have minimal access to all levels of linguistic processing (i.e., phonetic, phonological, lexical, suprasegmental) after adoption and after early exposure to English. However, the adoptees demonstrated an advantage at pitch perception and tone discrimination compared to non-tonal language speakers, suggesting that they may have some sensitivity to earlier acquired linguistic representations (i.e., phonology) with re-exposure to Mandarin, though this re-exposure does not offer an advantage for higher-level linguistic representations (i.e., lexical, suprasegmental) Overall, the behavioral data are consistent with the neural commitment theory that humans' language acquisition is attuned to their language environment early on in life. This study provides evidence that when there is a change in linguistic environment, an individual may replace previously learned linguistic information that does not align with their current linguistic environment.

REFERENCES

(n.d.). Geography and Population. Retrieved from: http://www.gcs.gov.mo/files/factsheet/geography.php?PageLang=E

Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2014). Rapid acquisition of auditory subcortical steady state responses using multichannel recordings. *Clinical Neurophysiology*, *125*(9), 1878-1888.

Boersma, P. & Weenink, D.(2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.39, retrieved 3 April 2018 from http://www.praat.org/

Carcagno, S., & Plack, C. J. (2011). Subcortical plasticity following perceptual learning in a pitch discrimination task. *Journal of the Association for Research in Otolaryngology*, *12*(1), 89-100.

Civil Services Bureau. (2017). *Official Languages Division.* Retrieved from: http://www.csb.gov.hk/english/aboutus/org/scsd/1470.html

Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*, *1*(2), 106-120.

Eimas, P. D., Siqueland, E. R., Juscyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303-306.

Gandour, J., Wong, D., & Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. *NeuroReport*, *9*(9), 2115-2119.

Hecox, K., & Galambos, R. (1974). Brain stem auditory evoked responses in human infants and adults. *Archives of Otolaryngology*, *99*(1), 30-33.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*(5), 393-402.

Hyltenstam, K., Bylund, E., Abrahamsson, N., & Park, H. S. (2009). Dominant-language replacement: The case of international adoptees. *Bilingualism: Language and Cognition*, *12*(2), 121-140.

International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

Isurin, L. (2000). Deserted island or a child's first language forgetting. *Bilingualism: Language and Cognition*, *3*(2), 151-166.

Jeng, F. C., Lin, C. D., Chou, M. S., Hollister, G. R., Sabol, J. T., Mayhugh, G. N., ... & Wang, C. Y. (2016). Development of subcortical pitch representation in three-month-old Chinese infants. *Perceptual and Motor Skills*, *122*(1), 123-135.

Jeng, F. C., Schnabel, E. A., Dickman, B. M., Hu, J., Li, X., Lin, C. D., & Chung, H. K. (2010). Early maturation of frequency-following responses to voice pitch in infants with normal hearing. *Perceptual and Motor Skills*, *111*(3), 765-784.

Jewett, D. L., & Williston, J. S. (1971). Auditory-evoked far fields averaged from the scalp of humans. *Brain*, *94*(4), 681.

Johnson, K. L., Nicol, T., Zecker, S. G., & Kraus, N. (2008). Developmental plasticity in the human auditory brainstem. *Journal of Neuroscience*, *28*(15), 4000-4007.

Klein, D., Zatorre, R. J., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage*, *13*(4), 646-653.

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: representation of pitch contours in Chinese tones. *Hearing Research*, *189*(1), 1-12.

Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language*, *110*(3), 135-148.

Krizman, J., Tierney, A., Fitzroy, A. B., Skoe, E., Amar, J., & Kraus, N. (2015). Continued maturation of auditory brainstem function during adolescence: A longitudinal approach. *Clinical Neurophysiology*, *126*(12), 2348-2355.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews. Neuroscience*, *5*(11), 831-843.

Kuhl, P. K., Tsao, F. M., Liu, H. M., Zhang, Y., & Boer, B. (2001). Language/culture/mind/ brain. *Annals of the New York Academy of Sciences*, *935*(1), 136-174.

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, *10*(3), 241-265.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, *106*(3), 1367-1381.

Moulines, E., & Charpentier, F. (1990). “Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones.” *Speech Communication*, *9*, 453-467.

Nicoladis, E., & Grabois, H. (2002). Learning English and losing Chinese: A case study of a child adopted from China. *International Journal of Bilingualism*, *6*(4), 441-454.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Pallier, C., Dehaene, S., Poline, J. B., LeBihan, D., Argenti, A. M., Dupoux, E., & Mehler, J. (2003). Brain imaging of language plasticity in adopted adults: Can a second language replace the first? *Cerebral Cortex*, *13*(2), 155-161.

Perrachione, T. K. (2014). *Pitch contour perception test (PCPT).* Retrieved from https://hdl.handle.net/2144/16461

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461-472.

Pierce, L. J., Klein, D., Chen, J. K., Delcenserie, A., & Genesee, F. (2014). Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(48), 17314-17319.

Reetzke, R., Xie, Z., Llanos, F., & Chandrasekaran, B. (2018). Tracing the Trajectory of Sensory Plasticity across Different Stages of Speech Learning in Adulthood. *Current Biology*.

Roberts, J. A., Pollock, K. E., Krakow, R., Price, J., Fulmer, K. C., & Wang, P. P. (2005). Language development in preschool-age children adopted from China. *Journal of Speech, Language, and Hearing Research*, *48*(1), 93-107.

Salamy, A. (1984). Maturation of the auditory brainstem response from birth through early childhood. *Journal of Clinical Neurophysiology*, *1*(3), 293-329.

Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex*, *25*(6), 1415-1426.

Snedeker, J., Geren, J., & Shafto, C. L. (2007). Starting over: International adoption as a natural experiment in language development. *Psychological Science*, *18*(1), 79-87.

Spapé, M. M., & Hommel, B. (2008). He said, she said: Episodic retrieval induces conflict adaptation in an auditory Stroop task. *Psychonomic Bulletin & Review*, *15*(6), 1117-1121.

Spitzer, E., White-Schwoch, T., Carr, K. W., Skoe, E., & Kraus, N. (2015). Continued maturation of the click-evoked auditory brainstem response in preschoolers. *Journal of the American Academy of Audiology*, *26*(1), 30-35.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *NeuroReport*, *19*(11), 1163.

Ventureyra, V. A., Pallier, C., & Yoo, H. Y. (2004). The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics*, *17*(1), 79-91.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*(5), 672.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Attention, Perception, & Psychophysics, 37*(1), 35-44.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63.

Whalen, D. H., Levitt, A. G., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French-and English-learning infants. *Journal of Child Language*, *18*(3), 501-516.

Worden, F. G., & Marsh, J. T. (1968). Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalography and Clinical Neurophysiology*, *25*(1), 42-52.

Wong, P. C., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. *Journal of Neuroscience*, *24*(41), 9153-9160.

Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *The Journal of the Acoustical Society of America*, *133*(1), 434-443.

Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics, 28*(4), 565-585.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*(4), 420-422.

Wong, P., & Strange, W. (2017). Phonetic complexity affects children's Mandarin tone production accuracy in disyllabic words: A perceptual study. *PLoS One, 12*(8), e0182337.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America, 120*(2), 1063-1074.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language, 68*(2), 123-139.

Yu, L., & Zhang, Y. (2018). Testing native language neural commitment at the brainstem level: A cross-linguistic investigation of the association between frequency-following response and speech perception. *Neuropsychologia, 109*, 140-148.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256*(5058), 846-849.

Zhou, W., & Boersma, M. (2014). Perception of birth language tone contrasts by adopted Chinese children. In *4th International Symposium on Tonal Aspects of Language* (pp. 63-66).

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]